

ON THE MULTIVARIATE ASYMPTOTIC DISTRIBUTION OF SEQUENTIAL CHI-SQUARE STATISTICS

JAMES H. STEIGER

UNIVERSITY OF BRITISH COLUMBIA

ALEXANDER SHAPIRO AND MICHAEL W. BROWNE

UNIVERSITY OF SOUTH AFRICA

The multivariate asymptotic distribution of sequential Chi-square test statistics is investigated. It is shown that: (a) when sequential Chi-square statistics are calculated for nested models on the same data, the statistics have an asymptotic intercorrelation which may be expressed in closed form, and which is, in many cases, quite high; and (b) sequential Chi-square *difference* tests are asymptotically independent. Some Monte Carlo evidence on the applicability of the theory is provided.

Key words: Asymptotic distribution theory, sequential Chi-square tests.

1. Introduction

In many situations in multivariate analysis one decides, using statistical information, which of several nested structural models appears to be “best” for a population of interest. Characteristically, a sample-based “discrepancy function” is evaluated for each nested model, and some statistical (or perhaps quasi-statistical) rationale is applied to the discrepancy function values. Ultimately, the procedures choose a model solely on the basis of discrepancy function values.

Consider, for example, the case of common factor analysis, in which one attempts to ascertain ℓ^* , the “best” value for the number of common factors ℓ , to retain in fitting a $p \times p$ covariance matrix. Perhaps the simplest and most popular approach to choosing ℓ^* is the sequential Chi-square test (SCT). In this procedure one chooses a starting value ℓ_1 (often $\ell_1 = 1$) for ℓ . The Chi-square “goodness of fit” test is performed in order to test the hypothesis that the factor model fits perfectly in the population for $\ell = \ell_1$. If the hypothesis is rejected at an α significance level, the significance test is repeated for $\ell_2 > \ell_1$ (often $\ell_2 = \ell_1 + 1$) common factors. (Note that the common factor model with ℓ_1 common factors is nested within the common factor model with ℓ_2 , $\ell_2 > \ell_1$, factors). The procedure terminates when, for some $\ell^* \geq \ell_1$, the null hypothesis of perfect fit is not rejected. The SCT method for choosing the “true” model has a number of flaws which have been described (Cliff, 1983; Cudeck & Browne, 1983) clearly in a form accessible to non-specialists.

An alternative approach is the sequential Chi-square difference test (SCDT). In this procedure, two nested models are compared by treating the difference of their Chi-square test statistics as a Chi-square statistic with degrees of freedom equal to the difference between the degrees of freedom for the individual Chi-squares. Here, the basic idea is to

This research was carried out while the first author was Visiting Professor in the Department of Statistics in the University of South Africa, and was supported in part by a research grant (NSERC #67-4640) from the National Sciences and Engineering Council of Canada to the first author. The support of both of these organizations is acknowledged with gratitude.

Please send requests for reprints to James H. Steiger, Department of Psychology, 2136 West Mall, University of British Columbia, Vancouver, B.C. CANADA V6T 1W5.

interpret the difference Chi-square as an "improvement index." So long as the difference Chi-square is significant, it indicates that a "significant improvement" in the fit of the model has been produced by changing the number of free parameters. The SCDT may also be repeated to compare more than two nested models.

A number of other rationales for utilizing the statistical information in a sequence of sample discrepancy functions have been proposed (Akaike, 1973; Tucker & Lewis, 1973; Schwartz, 1978; Steiger & Lind, 1980; Cudeck & Browne, 1983).

The methods described above use a sequence of discrepancy function values all calculated on the same observations. Hence one would not expect the test statistics to be independent. Evaluation of model selection methods and prediction of their performance might thus be enhanced if the multivariate distributional characteristics of the set of discrepancy function values were understood. In this paper we derive some basic results on the asymptotic multivariate distribution of sequential discrepancy function values (for nested models) calculated on the same data. We develop an asymptotic distributional characterization for the set of discrepancy function values which leads, in turn, to a simple expression for their asymptotic intercorrelation. Somewhat surprisingly, we find that there are numerous situations in which the correlation between SCT's will be very high whereas the SCDT's are, in fact, asymptotically independent.

Hogg (1961) has pointed out that, for some linear models, sequential likelihood ratio difference test statistics are independently distributed if all null hypotheses are true. Examples were given. Our results on sequential Chi-square difference tests differ from those of Hogg in that we examine asymptotic independence in the context of moment structures, and consider a general class of models, general discrepancy functions, and the situation where null hypotheses need not be true.

In section 2 we present our notation, regularity conditions, and background information. Section 3 contains the derivation of new results. Section 4 presents Monte Carlo evidence on the applicability of the results to common factor analysis, while section 5 gives the summary and conclusions.

2. Background, Notation, and Regularity Conditions

Let $\xi \in R^m$ be a vector variable which, in our applications, will be a parameter vector of some statistical population. For example, in the analysis of covariance structure models ξ will represent the nonduplicated elements of the $p \times p$ covariance matrix Σ , with $m = p(p + 1)/2$. A structural model M for ξ is an $m \times 1$ vector-valued function (mapping) $\mathbf{g}(\theta) = (g_1(\theta), \dots, g_m(\theta))'$, which relates the $q \times 1$ parameter vector θ , from a specified parameter space Ω , to ξ . The functions g_i , $i = 1, \dots, m$ are assumed to be continuously differentiable, and $\Omega \subseteq R^q$.

Consider two models, M_1 and M_2 . We say that M_2 is *nested within* M_1 , and denote this by $M_1 > M_2$, if both models involve the same mapping $\mathbf{g}(\theta)$, but the parameter space Ω_2 of M_2 is a subset of the parameter space Ω_1 of M_1 defined by the imposition of equality constraints. We consider a sequence $M_1 > M_2 > \dots > M_r$ of nested models with the mapping $\mathbf{g}(\theta)$ and the parameter spaces $\Omega_1, \dots, \Omega_r$ defined as follows. Ω_2 is a subset of Ω_1 given by imposing k_1 equality constraints, and Ω_3 is a subset of Ω_2 defined by adding k_2 equality constraints, that is,

$$\begin{aligned} \Omega_2 &= \{\theta \in \Omega_1 : c_i(\theta) = 0, i = 1, \dots, k_1\}, \\ \Omega_3 &= \{\theta \in \Omega_2 : c_i(\theta) = 0, i = k_1 + 1, \dots, k_1 + k_2\}, \quad \text{etc.} \end{aligned} \quad (1)$$

The constraint mapping $\mathbf{c}(\theta) = (c_1(\theta), \dots, c_k(\theta))'$, $k = k_1 + k_2 + \dots + k_{r-1}$, corresponds to the most constrained model M_r . The constraint functions $c_i(\theta)$ are assumed to be continuously differentiable. In many practical applications they are simple linear functions as-

signing specified values to elements of θ (e.g., $\theta_1 = 0$), or requiring equality of elements of θ (e.g., $\theta_1 = \theta_2$).

For a given sample estimate \hat{x}_n , based on a sample of size n , of the population value of ξ , one tries to fit the model

$$\xi = g(\theta) \tag{2}$$

by minimizing the discrepancy between \hat{x}_n and ξ , which is measured by means of a *discrepancy function* $F(\cdot, \cdot)$. Here $F(x, \xi)$ is a real-valued function of two vector variables $x, \xi \in R^m$ satisfying the following conditions (Browne, 1982, 1984; Shapiro, 1985):

- (i) $F(x, \xi) \geq 0$ for all x, ξ .
- (ii) $F(x, \xi) = 0$ if and only if $x = \xi$.
- (iii) F is twice continuously differentiable in x and ξ .

Then a sample estimate $\hat{\theta}_n^{(j)}, j = 1, \dots, r$, corresponding to the model M_j is chosen to minimize $F(\hat{x}_n, g(\cdot))$ over Ω_j , i.e.,

$$F(\hat{x}_n, g(\hat{\theta}_n^{(j)})) = \min_{\theta \in \Omega_j} F(\hat{x}_n, g(\theta)). \tag{3}$$

The corresponding minima will be denoted by $\hat{F}_n^{(j)}$,

$$\hat{F}_n^{(j)} = F(\hat{x}_n, g(\hat{\theta}_n^{(j)})), \quad j = 1, \dots, r$$

In practical applications it is usually not reasonable to assume that a model is a precise representation of reality. A more realistic view (Cudeck & Browne, 1983) is that a model is an approximation. We therefore allow for the situation where some (or even all) of the models M_j do not hold exactly in the sense that there is no θ from the parameter space such that $g(\theta)$ is equal to the population value of ξ . In order to ensure the existence of asymptotic distributions, we formulate a mathematical assumption which implies that the systematic errors in $\xi - g(\theta)$ are not "too large" relative to the stochastic errors in $\hat{x}_n - \xi$. That is, following Stroud (1972), we consider a *sequence* $\{\xi_n\}$ of population values of ξ converging to a point ξ_0 where all models hold, i.e., there exists $\theta_0 \in \Omega_r$ such that

$$\xi_0 = g(\theta_0).$$

Of course $\theta_0 \in \Omega_i, i = 1, \dots, r - 1$, as well.

We define the (population) *badness of fit* for models $M_j, j = 1, \dots, r$, as

$$\delta_n^{(j)} = \min_{\theta \in \Omega_j} F(\xi_n, g(\theta)). \tag{4}$$

Clearly $0 \leq \delta_n^{(1)} \leq \delta_n^{(2)} \leq \dots \leq \delta_n^{(r)}$.

We assume that $n^{1/2}(\hat{x}_n - \xi_n)$ has asymptotically a multivariate normal distribution with a null mean vector and a certain covariance matrix Γ . The following regularity conditions will be assumed in the derivation of our results:

- (R1) The parameter space Ω_1 is compact.
- (R2) The parameter vector θ is identified as θ_0 in Ω_1 , i.e., $g(\theta^*) = g(\theta_0)$ and $\theta^* \in \Omega_1$ implies that $\theta^* = \theta_0$.
- (R3) θ_0 is an interior point of Ω_1 .
- (R4) The $m \times q$ Jacobian matrix,

$$\Delta = \frac{\partial g}{\partial \theta'} \Big|_{\theta = \theta_0}$$

is of full column rank q .

(R5) The Hessian matrix

$$2V_0 = \frac{\partial^2 F}{\partial \xi \partial \xi'}$$

of F at the point (\mathbf{x}_0, ξ_0) , $\mathbf{x}_0 = \xi_0$, is nonsingular and $V_0 = \Gamma^{-1}$.

(R6) The $k \times q$ Jacobian matrix

$$\left. \frac{\partial \mathbf{c}}{\partial \theta'} \right|_{\theta = \theta_0}$$

is of full rank k .

(R7) $n\delta_n^{(j)}$, $j = 1, \dots, r$, converges to a finite number δ_j as $n \rightarrow \infty$.

Some additional remarks should clarify our regularity conditions and their implications. Identification of θ in Ω_1 (Condition R2) also implies identification in the restricted spaces $\Omega_2, \dots, \Omega_r$. Conditions (R4) and (R5) imply that the Hessian matrix

$$\frac{\partial^2 F}{\partial \theta \partial \theta'} = 2\Delta' V_0 \Delta$$

of $F(\mathbf{x}, \mathbf{g}(\theta))$ at (\mathbf{x}_0, θ_0) is nonsingular. This, together with Conditions (R1) and (R2), implies that $F(\mathbf{x}, \mathbf{g}(\cdot))$ has a unique minimizer for every \mathbf{x} in a neighborhood of $\mathbf{x}_0 = \xi_0$ (Shapiro, 1983, Theorem 4.1), i.e., θ is *conditionally identified* given $\mathbf{x} = \xi_n$ for sufficiently large n .

Consequently, $\hat{\theta}_n^{(j)}$ is a consistent estimator of the corresponding (population) minimizer of $F(\xi_n, \mathbf{g}(\cdot))$ over Ω_j (Shapiro, 1984). Often Condition (R1) does not hold in practical applications, where the parameter space is typically unbounded and hence is not compact. Fortunately this condition can be replaced by a certain condition of *boundedness* (Shapiro, 1984).

Conditions (i)–(iii) alone imply that the Hessian matrix $2V_0$ is nonnegative definite and, together with Condition (R5), that V_0 is positive definite. Also (see Shapiro, 1985)

$$2V_0 = \left. \frac{\partial^2 F}{\partial \mathbf{x} \partial \mathbf{x}'} \right|_{(\mathbf{x}, \xi) = (\mathbf{x}_0, \xi_0)}$$

Condition (R7) implies that $|\xi_n - \xi_0|$ is $O(n^{-1/2})$, i.e., ξ_n converges to ξ_0 at a rate $n^{-1/2}$ (Stroud, 1972; Kendall & Stuart, 1979, p. 247; Shapiro, 1983, p. 61; Browne, 1984, p. 66). Condition (R5) is sufficient to ensure that the minimum discrepancy function test statistics have asymptotic Chi-square distributions (Browne, 1974, 1984; Shapiro, 1983). Finally we note that $0 \leq \delta_1 \leq \delta_2 \leq \dots \leq \delta_r$.

3. Results

We prove our main results in this section. As a notational convenience, we prefix an expression with “L” to indicate an asymptotic result. For example, “LCor” might be used to denote an asymptotic correlation.

Theorem 1. Under regularity Conditions (R1)–(R7) and our stated assumptions:

- (i) The test statistics $n\hat{F}_n^{(1)}, n\hat{F}_n^{(2)}, \dots, n\hat{F}_n^{(r)}$ have asymptotic noncentral Chi-square distributions with $v_1 = m - q$, $v_2 = m - q + k_1, \dots, v_r = m - q + k_1 + \dots + k_{r-1}$ degrees of freedom, and noncentrality parameters $\delta_1, \delta_2, \dots, \delta_r$, respectively.
- (ii) The statistics $n\hat{F}_n^{(1)}, n\hat{F}_n^{(2)} - n\hat{F}_n^{(1)}, \dots, n\hat{F}_n^{(r)} - n\hat{F}_n^{(r-1)}$ are mutually asymptotically independent and have asymptotic noncentral Chi-square distributions with $v_1 =$

$m - q, v_1 - v_2 = k_1, \dots, v_r - v_{r-1} = k_{r-1}$ degrees of freedom and noncentrality parameters $\delta_1, \delta_2 - \delta_1, \dots, \delta_r - \delta_{r-1}$, respectively.

Proof. It can be shown under conditions (R1) and (R2) that a minimizer of $F(\mathbf{x}, \mathbf{g}(\cdot))$ over Ω_j converges to θ_0 as \mathbf{x} tends to $\mathbf{x}_0 = \xi_0$ (Shapiro, 1984). We have that $\hat{\mathbf{x}}_n - \xi_n \xrightarrow{P} \mathbf{0}$, and, by (R7), $\xi_n \rightarrow \xi_0$. Consequently, $\hat{\mathbf{x}}_n \xrightarrow{P} \xi_0$, and hence $\hat{\theta}_n^{(j)} \xrightarrow{P} \theta_0$.

Because of Conditions (R3) and (R6), there exists a continuously differentiable one-to-one mapping

$$\theta = \mathbf{h}(\gamma)$$

from a neighborhood of a point $\gamma_0 \in R^q, \mathbf{h}(\gamma_0) = \theta_0$, to a neighborhood of θ_0 , such that $\gamma_i = c_i(\mathbf{h}(\gamma)), i = 1, \dots, k$. (Local Linearization Theorem, see, e.g., Klingenberg, 1978, Theorem 0.5.2, p. 6). Thus the equations $\mathbf{c}(\theta) = \mathbf{0}$ can be reduced (locally) to the simple case

$$\mathbf{c}(\gamma) = (\gamma_1, \dots, \gamma_k)' = \mathbf{0}$$

by a suitable reparameterization $\theta = \mathbf{h}(\gamma)$. Since $\mathbf{h}(\gamma)$ is one-to-one, the minimizations with respect to γ will be equivalent. (Note that, since the estimators $\hat{\theta}_n^{(j)}$ converge in probability to θ_0 , for large enough n it is sufficient to minimize $\hat{F}(\mathbf{x}_n, \mathbf{g}(\cdot))$ in a given neighborhood of θ_0 .) Therefore, in what follows, we can suppose without loss of generality that the nested model $M_j, j = 2, \dots, r$, is defined by $(\theta'_1, \dots, \theta'_{j-1}) = \mathbf{0}$, where $\theta' = (\theta'_1, \dots, \theta'_{r-1}, \theta'_r)$, where θ_i is $k_i \times 1, i = 1, \dots, r - 1$ and θ_r is $(q - k) \times 1$. Then M_j may be considered as an unrestricted model with respect to the parameter vector $(\theta'_j, \dots, \theta'_r)$. The corresponding Jacobian matrix is $(\Delta_j, \dots, \Delta_r)$, where

$$\Delta_i = \left. \frac{\partial \mathbf{g}}{\partial \theta'_i} \right|_{\theta = \theta_0}$$

is $m \times k_i, i = 1, \dots, r - 1$ and Δ_r is $m \times (q - k)$.

Under regularity Conditions (R1)–(R5), it can be shown (Shapiro, 1985) that

$$n\hat{F}_n^{(j)} \stackrel{a}{=} [n^{1/2}(\hat{\mathbf{x}}_n - \xi_0)]' H_j [n^{1/2}(\hat{\mathbf{x}}_n - \xi_0)], \tag{5}$$

where

$$H_j = \Phi_j(\Phi'_j V_0^{-1} \Phi_j)^{-1} \Phi'_j. \tag{6}$$

In (5) “ $\stackrel{a}{=}$ ” stands for “is asymptotically equal to” and means that the difference between the left and right sides approaches zero in probability as n tends to infinity.

In (6) Φ_j is an orthogonal complement of the Jacobian matrix $[\Delta_j, \dots, \Delta_r]$, i.e., Φ_j is an $m \times (m - q + k_1 + \dots + k_{j-1})$ matrix of full column rank such that $\Phi'_j[\Delta_j, \dots, \Delta_r] = \mathbf{0}$. Recalling (R5) we have

$$H_j = \Phi_j(\Phi'_j \Gamma \Phi_j)^{-1} \Phi'_j. \tag{7}$$

From (5) and (7) it follows that

$$n\hat{F}_n^{(j)} \stackrel{a}{=} \mathbf{z}'_j \mathbf{z}_j, \tag{8}$$

where

$$\mathbf{z}_j = (\Phi'_j \Gamma \Phi_j)^{-1/2} \Phi'_j [n^{1/2}(\hat{\mathbf{x}}_n - \xi_0)]. \tag{9}$$

It can be seen that \mathbf{z}_j has an asymptotic multivariate normal distribution with the identity covariance matrix and mean vector μ_j given by

$$\mu_j = (\Phi'_j \Gamma \Phi_j)^{-1/2} \Phi'_j [n^{1/2}(\xi_n - \xi_0)].$$

Therefore $n\hat{F}_n^{(j)}$ has an asymptotic noncentral Chi-square distribution with $v_j = m - q + k_1 + \dots + k_{j-1}$ degrees of freedom, and noncentrality parameter

$$\mu'_j \mu_j = n(\xi_n - \xi_0)' H_j (\xi_n - \xi_0). \tag{10}$$

It can be shown that $2H_j$ is the Hessian matrix of the min-function

$$F_{\min}^{(j)}(\mathbf{x}) = \min_{\boldsymbol{\theta} \in \Omega_j} F(\mathbf{x}, \mathbf{g}(\boldsymbol{\theta}))$$

at the point $\mathbf{x}_0 = \xi_0$ (Shapiro, 1985). Note that $F_{\min}^{(j)}(\mathbf{x}_0) = 0$ and that the first order partial derivatives of $F_{\min}^{(j)}$ at \mathbf{x}_0 vanish. Therefore, by Taylor's theorem, we have

$$nF_{\min}^{(j)}(\xi_n) = n(\xi_n - \xi_0)' H_j^* (\xi_n - \xi_0), \tag{11}$$

where the symmetric matrix H_j^* tends to H_j as $n \rightarrow \infty$. From (10) and (11) we obtain that as n increases the noncentrality parameter becomes

$$\lim_{n \rightarrow \infty} nF_{\min}^{(j)}(\xi_n) = \lim_{n \rightarrow \infty} n\delta_n^{(j)} = \delta_j.$$

This completes the proof of (i).

Now recall expressions (5)–(8). Let Φ_1 be a given $m \times (m - q)$ orthogonal complement of $\Delta = [\Delta_1, \dots, \Delta_r]$. We construct an orthogonal complement Φ_2 corresponding to M_2 as follows. Let Ψ be an $m \times k_1$ orthogonal complement of the $m \times (m - k_1)$ matrix $[\Gamma\Phi_1, \Delta_2, \dots, \Delta_r]$. Clearly $\Psi'[\Delta_2, \dots, \Delta_r] = 0$ and, since Φ_1 is an orthogonal complement of Δ , $\Phi_1' [\Delta_2, \dots, \Delta_r] = 0$. Therefore, Φ_2 , defined as

$$\Phi_2 = [\Psi, \Phi_1]$$

is an orthogonal complement corresponding to M_2 . Moreover, by the definition of Ψ , we have that

$$\Psi' \Gamma \Phi_1 = 0. \tag{12}$$

It follows from (12) that

$$\Phi_2' \Gamma \Phi_2 = \begin{bmatrix} \Psi' \Gamma \Psi & 0 \\ 0 & \Phi_1' \Gamma \Phi_1 \end{bmatrix}, \tag{13}$$

and hence (see (8))

$$\mathbf{z}'_2 = (\mathbf{w}', \mathbf{z}'_1),$$

where

$$\mathbf{w} = (\Psi' \Gamma \Psi)^{-1/2} \Psi' [n^{1/2}(\hat{\mathbf{x}}_n - \xi_0)].$$

It can be shown in a way very similar to the method used in the proof of part (i) that $\mathbf{w}'\mathbf{w}$ has an asymptotic noncentral Chi-square distribution with $v_2 - v_1 = k_1$ degrees of freedom and noncentrality parameter $\delta_2 - \delta_1$. Moreover, \mathbf{w} and \mathbf{z}_1 are asymptotically uncorrelated, since their asymptotic covariance matrix is

$$\text{LCov}(\mathbf{w}, \mathbf{z}_1) = (\Psi' \Gamma \Psi)^{-1/2} \Psi' \Gamma \Phi_1 (\Phi_1' \Gamma \Phi_1)^{-1/2} = 0,$$

and thus they are asymptotically independent. This implies that $\mathbf{z}'_1 \mathbf{z}_1$ and $\mathbf{w}'\mathbf{w}$ are asymptotically independent. Noting that

$$n\hat{F}_n^{(2)} - n\hat{F}_n^{(1)} = \mathbf{z}'_2 \mathbf{z}_2 - \mathbf{z}'_1 \mathbf{z}_1 = \mathbf{w}'\mathbf{w}$$

we obtain that $n\hat{F}_n^{(2)} - n\hat{F}_n^{(1)}$ has an asymptotic noncentral Chi-square distribution with $v_2 - v_1$ degrees of freedom and noncentrality parameter $\delta_2 - \delta_1$, and that $n\hat{F}_n^{(2)}$ and $n\hat{F}_n^{(1)}$

– $n\hat{F}_n^{(1)}$ are asymptotically independent. The proof may be completed for the other difference statistics in a similar manner. \square

Theorem 1 leads to the following closed form expression for the asymptotic correlation matrix of Chi-square statistics.

Theorem 2. Let $M_1 > M_2 > \dots > M_r$ be a sequence of nested models and $n\hat{F}_n^{(1)}, n\hat{F}_n^{(2)}, \dots, n\hat{F}_n^{(r)}$ be the corresponding test statistics. Then the asymptotic correlation between $n\hat{F}_n^{(i)}$ and $n\hat{F}_n^{(j)}, j > i$, is

$$\left[\frac{v_i + 2\delta_i}{v_j + 2\delta_j} \right]^{1/2} \tag{14}$$

Proof.

$$\begin{aligned} LCov \{n\hat{F}_n^{(i)}, n\hat{F}_n^{(j)}\} &= LCov \{n\hat{F}_n^{(i)}, (n\hat{F}_n^{(j)} - n\hat{F}_n^{(i)}) + n\hat{F}_n^{(i)}\}, \\ &= LCov \{n\hat{F}_n^{(i)}, n\hat{F}_n^{(j)} - n\hat{F}_n^{(i)}\} + LCov \{n\hat{F}_n^{(i)}, n\hat{F}_n^{(i)}\}, \\ &= 0 + L \text{Var} \{n\hat{F}_n^{(i)}\}. \end{aligned}$$

Now

$$\begin{aligned} LCor \{n\hat{F}_n^{(i)}, n\hat{F}_n^{(j)}\} &= \frac{LCov \{n\hat{F}_n^{(i)}, n\hat{F}_n^{(j)}\}}{[LVar \{n\hat{F}_n^{(i)}\}LVar \{n\hat{F}_n^{(j)}\}]^{1/2}}, \\ &= \left[\frac{LVar \{n\hat{F}_n^{(i)}\}}{LVar \{n\hat{F}_n^{(j)}\}} \right]^{1/2} \end{aligned} \tag{15}$$

The variance of a noncentral Chi-square variable with v degrees of freedom and noncentrality parameter δ is $2v + 4\delta$ (Johnson & Kotz, 1970, p. 134). Substituting this expression in (15), we obtain (14). \square

It is of interest to note that this asymptotic correlation matrix has Guttman’s (1954) perfect simplex structure.

4. Some Monte Carlo Results

The results presented in section 3 are based on asymptotic distribution theory which may yield a poor approximation if n is not large. Our derivations also relied on an assumption that population values converge at the rate of $n^{-1/2}$ to a point where both models are correct. In practice, this indicates that the noncentral Chi-square approximation will be reasonably effective so long as the noncentrality parameter is not “too large.”

The present results, however, are especially convenient because of their simplicity. Ultimately, what must be determined are (a) whether adequate convergence to the asymptotic result occurs at realistic sample sizes, and (b) whether the results of section 3 provide an adequate approximation for realistically large values of the noncentrality parameters $n\delta_n^{(j)}$.

To answer these questions in the broad context of possible applications of our results is probably not feasible, and authoritative answers, even in the context of a simple application (such as common factor analysis) would require an extensive (and very expensive) Monte Carlo study. Such an investigation is planned, but here we present some rather limited results using the factor analysis model. The factor patterns used were chosen deliberately to minimize problems with Heywood cases and non-convergence, to avoid con-

TABLE 1
Monte Carlo Conditions

General Form of the Population Factor Pattern

a	0	0
b	0	0
c	0	0
0	d	0
0	e	0
0	f	0
0	0	g
0	0	h
0	0	i

Condition	n	a	b	c	d	e	f	g	h	i
1	100	.447	.548	.633	.447	.548	.633	.447	.548	.633
2	200	.447	.548	.633	.447	.548	.633	.447	.548	.633
3	300	.447	.548	.633	.447	.548	.633	.447	.548	.633
4	400	.447	.548	.633	.447	.548	.633	.447	.548	.633
5	500	.447	.548	.633	.447	.548	.633	.447	.548	.633
6	600	.447	.548	.633	.447	.548	.633	.447	.548	.633
7	700	.447	.548	.633	.447	.548	.633	.447	.548	.633
8	800	.447	.548	.633	.447	.548	.633	.447	.548	.633
9	900	.447	.548	.633	.447	.548	.633	.447	.548	.633
10	1000	.447	.548	.633	.447	.548	.633	.447	.548	.633
11	500	.900	.800	.700	.900	.800	.700	.900	.800	.700
12	500	.300	.400	.500	.300	.400	.500	.300	.400	.500
13	500	.900	.900	.900	.700	.700	.700	.500	.500	.500
14	500	.900	.700	.500	.900	.700	.500	.900	.700	.500

founding questions of algorithmic performance with questions about the applicability of our statistical results.

There were 14 experimental conditions. In each, the population correlation matrix conformed exactly to the common factor model with 9 variables and 3 uncorrelated common factors. The factor pattern which generated the population correlation matrix was always of the same general form (having good "simple structure"), as shown in Table 1. To assess convergence properties, Conditions 1 through 10 examined the identical factor pattern under differing sample sizes which varied from 100 to 1000 in increments of 100. The remaining conditions examined several different factor patterns, all at a sample size of 500. Simulated sampling of correlation matrices for the multivariate normal distribution was accomplished by means of the Bartlett decomposition (e.g., Browne, 1968a). In implementing this method, normal random numbers were generated by the method of Kinderman and Ramage (1976), and Chi-square random variables by the method of Fishman (1976). Both algorithms are described in detail by Kennedy and Gentle (1980). There

TABLE 2

Comparison of Obtained and Predicted Test Statistic Intercorrelations

Condition	Test Statistic Intercorrelations			Difference Statistic Intercorrelations			
	ρ_{21}	ρ_{31}	ρ_{32}	$\rho_{1-2,2}$	$\rho_{1-2,3}$	$\rho_{2-3,3}$	$\rho_{1-2,2-3}$
1	.75 (.74)	.40 (.32)	.53 (.43)	.10	.06	.12	.09
2	.74 (.73)	.26 (.24)	.36 (.33)	.08	.03	.07	.08
3	.71 (.72)	.23 (.20)	.28 (.27)	-.05	.04	.02	-.07
4	.72 (.72)	.18 (.17)	.20 (.24)	-.00	.05	-.03	-.02
5	.70 (.72)	.13 (.15)	.18 (.22)	-.02	-.00	-.03	-.01
6	.71 (.71)	.17 (.14)	.20 (.20)	-.04	.03	.01	-.05
7	.70 (.71)	.17 (.13)	.20 (.18)	-.06	.03	.03	-.06
8	.74 (.71)	.14 (.12)	.17 (.17)	.01	.01	.01	.00
9	.70 (.71)	.11 (.12)	.14 (.16)	-.02	.02	-.02	-.02
10	.70 (.71)	.11 (.12)	.18 (.16)	.01	-.04	.03	.02
11	.71 (.71)	.11 (.07)	.09 (.10)	-.02	.06	-.01	-.03
12	.75 (.73)	.33 (.27)	.41 (.36)	.09	.06	.06	.07
13	.47 (.47)	.14 (.12)	.27 (.25)	.02	.02	.03	.02
14	.71 (.71)	.14 (.09)	.13 (.13)	-.03	.06	-.00	-.04

were 1000 Monte Carlo repetitions in each condition. Each generated sample correlation matrix was factor analysed three times (assuming 1, 2, and 3 common factors) by the method of maximum likelihood, using a Gauss-Seidel algorithm (Browne, 1968b). The Chi-square statistics were calculated using the Bartlett correction factor as n^*F , where $n^* = n - 29/6 - 2\ell/3$, ℓ is the fitted number of common factors, and F the maximum likelihood discrepancy function. Results are summarized in Tables 2 and 3.

Table 2 compares the obtained and predicted test statistic intercorrelations for the 14 experimental conditions. Predicted correlations are given in parentheses. The left side of Table 2 gives the obtained and predicted correlations between the Chi-square goodness-of-fit statistics for 1, 2, and 3 common factor models. For example, ρ_{21} refers to the correlation between the Chi-square goodness of fit statistics for 1-factor and 2-factor common factor models. The right side of the table gives correlations for Chi-square difference statistics. For example, $\rho_{1-2,2}$ refers to the correlation between "1-2" (the difference between the Chi-square statistics for 1 and 2 common factors), and "2" (the Chi-square statistic for 2 common factors). The theory developed in this paper predicts that all of these correlations are zero.

The data in Table 2 indicate that, in general, differences between obtained and predicted test statistic intercorrelations are trivial for sample sizes above 400.

Table 3 compares the obtained and predicted means and variances of the chi-square statistics in the 14 experimental conditions. (For compactness, results for the noncentral case are given in scientific notation, with mantissa for the predicted values shown in parentheses. In the central 3-factor case, the predicted value of the mean was always 12, and the predicted value of the variance was always 24.) Here, the general trends in the

TABLE 3

Comparison of Obtained and Predicted Test Statistic Means and Variances

Condition	Means			Variances		
	1	2	3	1	2	3
1	6.60E1(7.27)	3.43E1(4.20)	11.1	2.39E2(2.37)	1.13E2(1.30)	20.3
2	1.12E2(1.21)	5.79E1(6.59)	11.6	4.82E2(4.28)	2.28E2(2.26)	20.1
3	1.61E2(1.68)	8.26E1(8.98)	11.6	6.63E2(6.19)	3.67E2(3.21)	24.4
4	2.09E2(2.16)	1.07E2(1.14)	11.8	8.90E2(8.10)	4.70E2(4.33)	24.6
5	2.59E2(2.64)	1.32E2(1.38)	11.8	1.14E3(1.00)	5.78E2(5.12)	24.7
6	3.08E2(3.11)	1.57E2(1.61)	12.0	1.24E3(1.19)	6.69E2(6.07)	23.7
7	3.53E2(3.59)	1.82E2(1.85)	11.8	1.44E3(1.38)	7.93E2(7.03)	23.3
8	4.05E2(4.07)	2.08E2(2.09)	12.2	1.85E3(1.57)	1.01E3(0.80)	25.8
9	4.52E2(4.55)	2.31E2(2.33)	12.1	1.92E3(1.76)	9.86E2(8.94)	24.4
10	5.01E2(5.02)	2.56E2(2.57)	11.9	2.21E3(1.96)	1.06E3(0.99)	24.6
11	1.29E3(1.29)	6.54E2(6.53)	12.3	4.82E3(5.12)	2.54E3(2.57)	28.3
12	8.96E1(9.86)	4.58E1(5.49)	11.2	3.16E2(3.41)	1.50E2(1.82)	19.2
13	4.40E2(4.40)	1.03E2(1.03)	11.9	1.94E3(1.71)	3.93E2(3.75)	22.7
14	7.63E2(7.62)	3.88E2(3.87)	12.2	2.82E3(2.99)	1.51E3(1.51)	28.1

data seem to be (a) that convergence is faster in the central case than in the non-central case. Convergence to predicted values seems reasonably good for $n = 500$ or higher, although this seems to vary with different population factor patterns.

Overall, the Monte Carlo evidence seems to support the notion that Theorems 1 and 2 yield accurate predictions of the moment structure of the SCT and the SCDT, provided that the sample size is reasonably large.

5. Summary and Conclusions

Theorems 1 and 2 establish results on the joint asymptotic distributions of Chi-square difference tests, on a sequence of nested models. It should be emphasized that these results are quite general—they are not restricted to maximum likelihood and generalized least squares discrepancy functions, but hold for any discrepancy function satisfying conditions (i), (ii), (iii), and regularity condition (R5).

The Monte Carlo evidence suggests that the predictions of the theory are essentially correct in the case of maximum likelihood common factor analysis. Future Monte Carlo investigations should examine (a) the effectiveness of the theory for other models and discrepancy functions, and (b) robustness of its predictions to violations of assumptions and regularity conditions.

References

- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & F. Csaki (Eds.), *Second International Symposium on Information Theory*. Budapest: Akademiai Kiado.
- Browne, M. W. (1968a). A comparison of factor analytic techniques. *Psychometrika*, 33, 267–334.
- Browne, M. W. (1968b). Gauss-Seidel computing procedures for a family of factor analytic solutions (Research Bulletin RB-68-61). Princeton, N.J.: Educational Testing Service.

- Browne, M. W. (1974). Generalized least squares in the analysis of covariance structures. *South African Statistical Journal*, 8, 1-24.
- Browne, M. W. (1982). Covariance structures. In D. M. Hawkins (Ed.), *Topics in applied multivariate analysis* (pp. 72-141). Cambridge: Cambridge University Press.
- Browne, M. W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, 37, 62-83.
- Cliff, N. (1983). Some cautions concerning the application of causal modeling methods. *Multivariate Behavioral Research*, 18, 115-126.
- Cudeck, R., & Browne, M. W. (1983). Cross-validation of covariance structures. *Multivariate Behavioral Research*, 18, 147-167.
- Fishman, G. S. (1976). Sampling from the gamma distribution on a computer. *Communications of the Association of Computing Machinery*, 19, 407-409.
- Guttman, L. (1954). A new approach to factor analysis: The radex. In P. F. Lazarsfeld (Ed.), *Mathematical thinking in the social sciences* (pp. 216-348). Glencoe, IL: Free Press.
- Hogg, R. V. (1961). On the resolution of statistical hypotheses. *Journal of the American Statistical Association*, 56, 978-989.
- Johnson, N. L., & Kotz, S. (1970). *Continuous univariate distributions—2*. Boston: Houghton-Mifflin.
- Kendall, M. G., & Stuart, A. (1979). *The advanced theory of statistics: Vol. 2* (4th ed.) London: Griffin.
- Kennedy, W. J., & Gentle, J. E. (1980). *Statistical computing*. New York: Marcel Dekker.
- Kinderman, A. J., & Ramage, J. G. (1976). Computer generation of normal random variables. *Journal of the American Statistical Association*, 71, 893-896.
- Klingenberg, W. (1978). *A course in differential geometry*. New York: Springer-Verlag.
- Schwartz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6, 461-464.
- Shapiro, A. (1983a). Asymptotic distribution theory in the analysis of covariance structures (a unified approach). *South African Statistical Journal*, 17, 33-81.
- Shapiro, A. (1984). A note on the consistency of estimators in the analysis of moment structures. *British Journal of Mathematical and Statistical Psychology*, 37, 84-88.
- Shapiro, A. (1985). Asymptotic equivalence of minimum discrepancy function estimators to G.L.S. estimators. *South African Statistical Journal*, 19, 73-81.
- Steiger, J. H., & Lind, J. C. (1980). Statistically-based tests for the number of common factors. Paper presented at the annual meeting of the Psychometric Society, Iowa City, IA.
- Stroud, T. W. (1972). Fixed alternatives and Wald's formulation of the noncentral asymptotic behavior of the likelihood ratio statistic. *Annals of Mathematical Statistics*, 43, 447-454.
- Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, 38, 1-10.

Manuscript received 1/5/84

Final version received 2/12/85